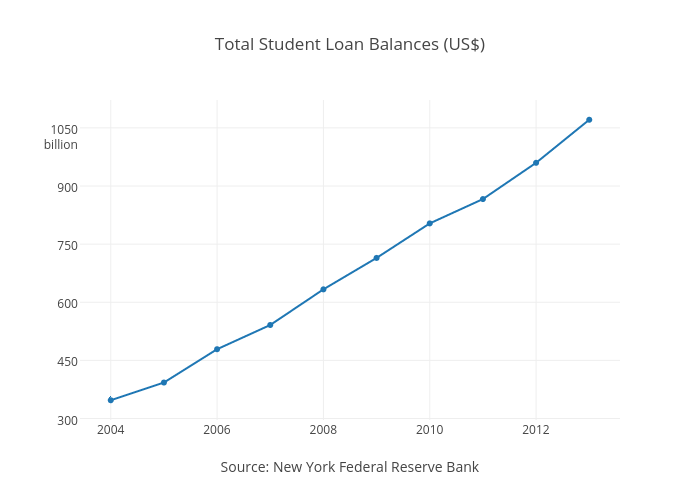
* <https://medium.com/@plotlygraphs/how-to-analyze-data-eight-useful-ways-you-can-make-graphs-328b81c2fa13>
* <https://towardsdatascience.com/10-viz-every-ds-should-know-4e4118f26fc3>
* <https://www150.statcan.gc.ca/n1/edu/power-pouvoir/ch9/pie-secteurs/5214826-eng.htm>
* <https://www.thoughtco.com/what-are-time-series-graphs-3126233>
* <https://towardsdatascience.com/the-complete-guide-to-time-series-analysis-and-forecasting-70d476bfe775> (Time series graph)

**In previous section we discussed Two important chart, Bar Chart and Pie Chart. In this session we shall discuss other two charts**

* **Line Chart**
* **Area chart**
* **Time series line chart**

1. **Line chart**

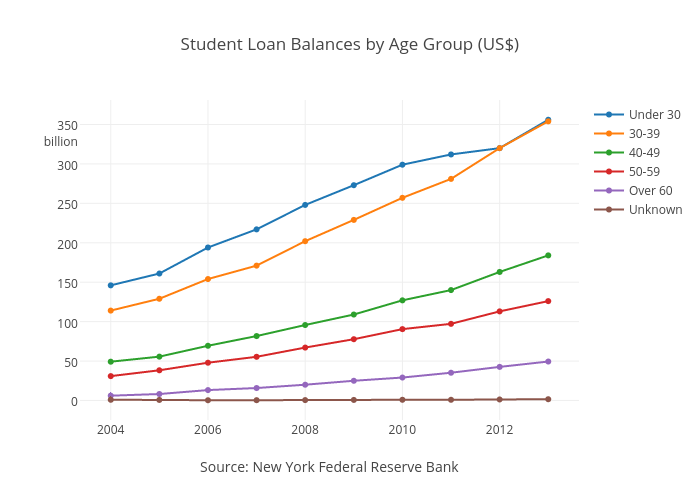
**Line charts** are similar to bar charts in that they compare values, but the x-axis displays continuous, rather than discrete, data. Each data point has an x and y value and is connected by a line. This basic line chart shows the total American student **loan debt**over time, with the year along the x-axis, and the amount of debt along the y-axis. Line charts emphasize the overall trend or pattern of the data. (You can draw Bar Chart for the same line chart)



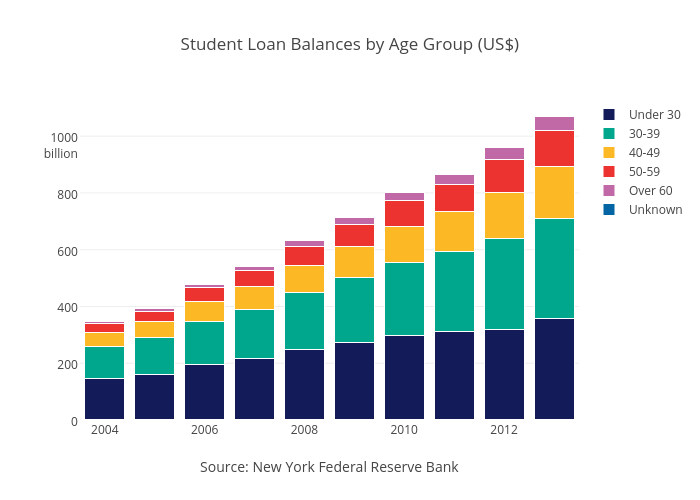
**Variation in Line Plotting:**

For Better visualization and understanding of Data, the simple line plot can be redrawn

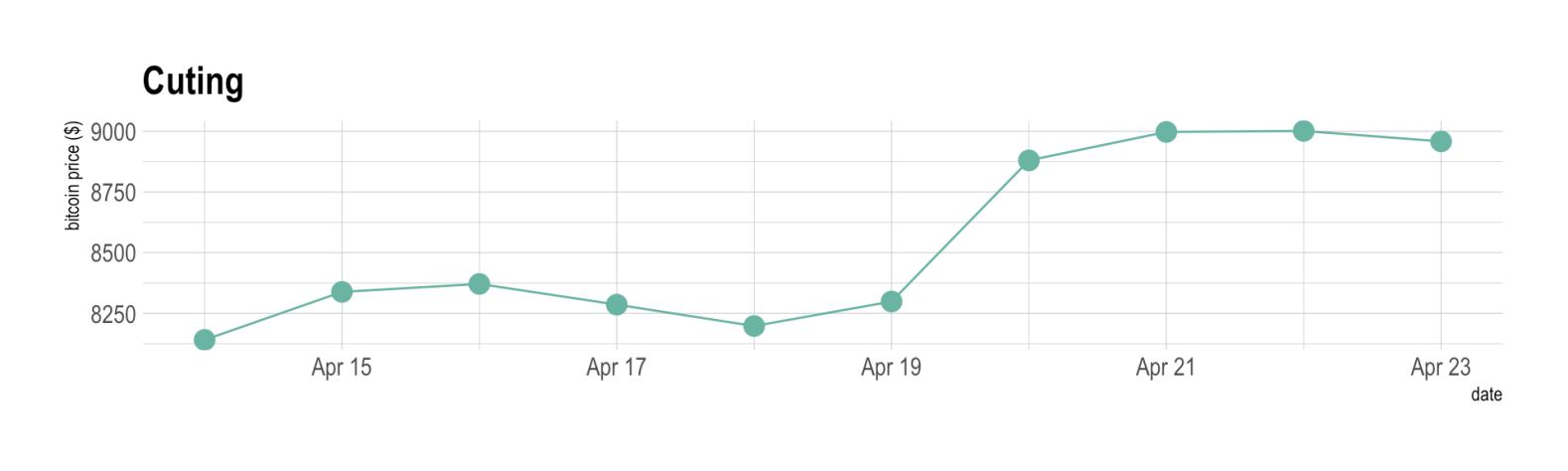
1. You can compare multiple traces on a line chart. Following graph is made from the same data, but broken down into age group.



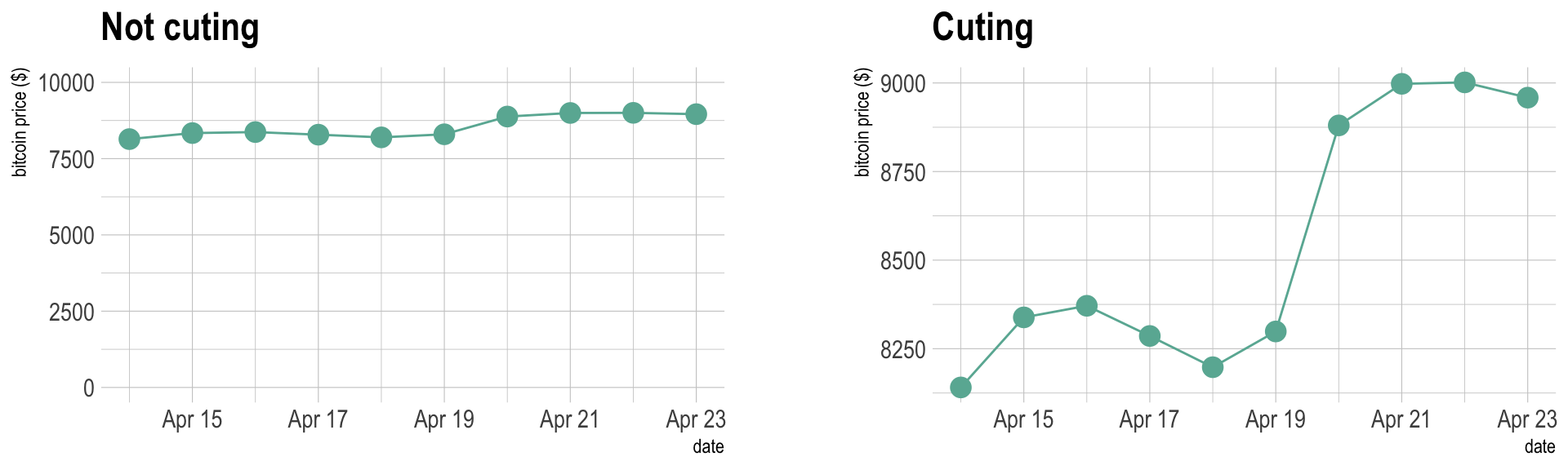
The same data is plotted as a bar chart in next figure. This is a more effective example of a stacked bar chart, as the total debt for all age groups can be compared between years.



1. If the number of data points is low, it is advised to represent each individual observation with a dot. It allows to understand when exactly the observation have been made:



1. Cut the Y axis: Weather or not the Y axis must start at 0 is a hot topic leading to intense debates. The graphic below presents the same data, starting at 0 (left) or not (right). Generally, line plot do not need to start at 0 since it allows to observe patterns more efficiently.

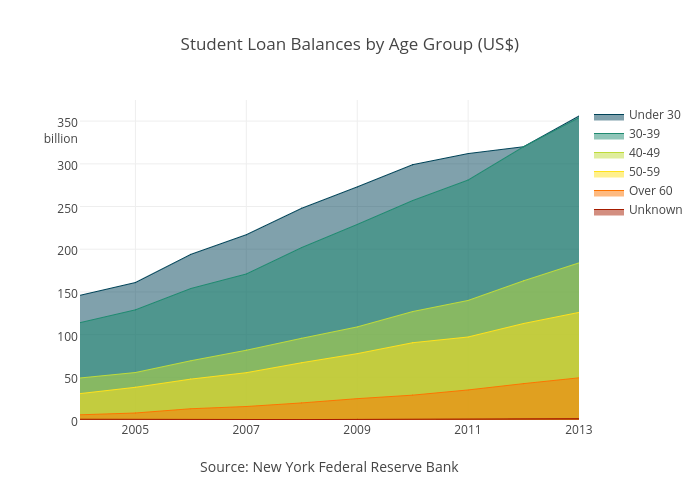


Note :

* It can be tricky to choose between a line and bar chart **when comparing values over time, as time** (Time Chart, Next section) can be represented as **continuous or discrete**.
* The stacked bar chart makes it easier to see the total amount of debt, while the line chart shows the trend for each age group separately.
* Line charts excel when the changes are more subtle or vary a lot over time.
* If you need to compare the evolution of 2 different variables, do not use [dual axis](https://www.data-to-viz.com/caveat/dual_axis.html). Indeed dual axis can show very different results depending on what range you apply to the axis. [Read more about it](https://www.data-to-viz.com/caveat/dual_axis.html).
* Mind the **spaghetti chart:** too many lines make the chart unreadable.
* Think about the [aspect ratio](https://www.data-to-viz.com/caveat/aspect_ratio.html) of the graphic, extreme ratio make the chart unreadable.

1. **Area Chart**

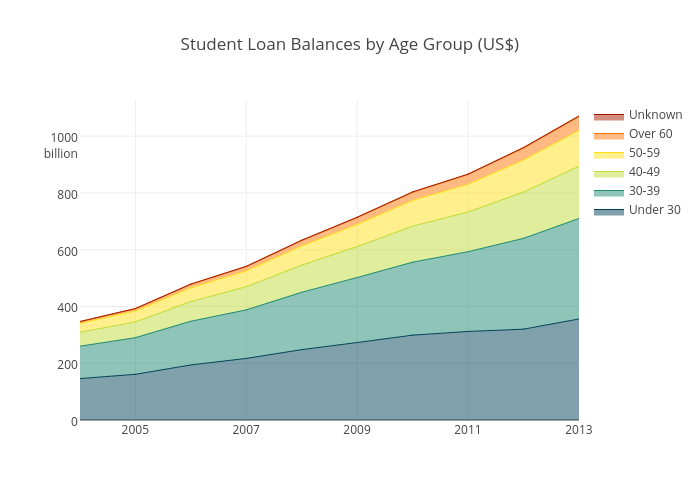
[**Area charts**](https://plot.ly/online-graphing/tutorials/how-to-make-an-area-graph/) are very similar to line charts, but the area under the line is filled in. This simple visual difference can be an effective way of drawing attention to the magnitude of difference between traces, or the cumulative value over a period of time. However, it can be confusing if there are many overlapping traces.



Here, all the information from the multiple line chart of previous section is preserved, but use of color draws attention to the volume of debt for each group. In this overlaid area graph, the areas are translucent and placed in front of each other.

**Variation:**

A stacked area graph adds each trace on top of the last one, such that the areas will never overlap.

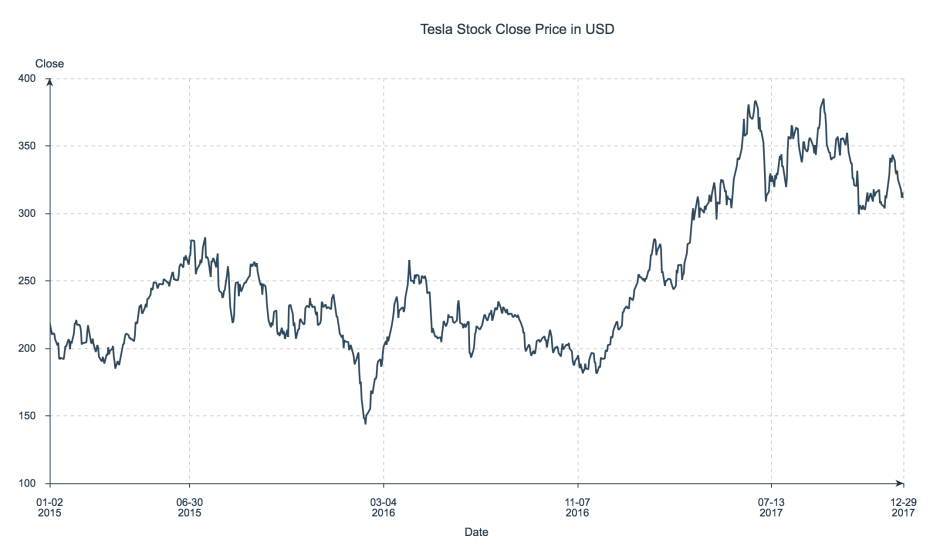


**Note :**

* Stacked area graphs have the benefit of showing the total of all the groups, but make it difficult to see values and patterns for the individual traces. Switch to stacked area charts by making sure your data is cumulative and then hit ‘fill to next y’ on the mode tab under traces. Fill to Y=0 makes an overlaid area chart.
* stacked area graph are appropriate to study the evolution of the whole and the relative proportions of each group. Indeed, the top of the areas allows to visualize how the whole behaves, like for a classic area chart.
* however they are not appropriate to study the evolution of each individual group: it is very hard to subtract the height of other groups at each time point. For a more accurate but less attractive figure, consider a [line chart](https://www.data-to-viz.com/graph/line.html) or [area chart](https://www.data-to-viz.com/graph/area.html) using small multiple.

1. **Time Series chart (time Plot/Time Series Graph)**

One feature of data that you may want to consider is that of time. A [graph](https://www.thoughtco.com/frequently-used-statistics-graphs-4158380) that recognizes this ordering and displays the change of the values of a variable as time progresses is called a time series graph. It is simply a scatter plot ir a line plot with a time range on the x-axis where each dot forms part of a line. Important to note that time is continuous, so this is a continuous graph (Not discrete).



**Time Series Plot of Tesla Stock Close Price from 2015–2017**

Time series plots are great for visually investigating trends, jumps and dumps in data over time, which makes them especially popular for financial and sensor data.

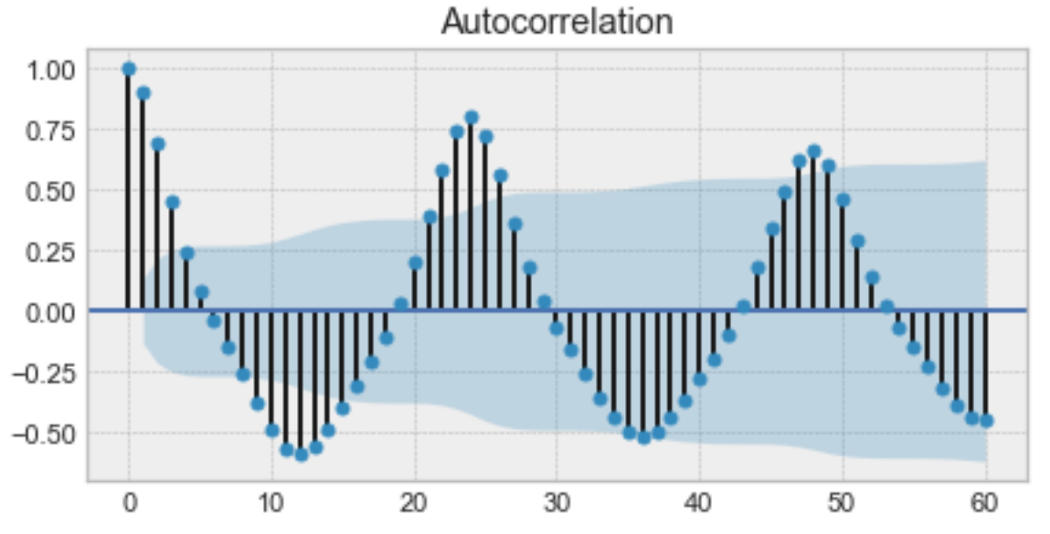
E.g. Suppose that you want to study the climate of a region for an entire month. Every day at noon you note the temperature and write this down in a log. A variety of statistical studies could be done with this data. You could find the [mean](https://www.thoughtco.com/the-mean-median-and-mode-2312604) or the [median](https://www.thoughtco.com/what-is-the-median-3126370) temperature for the month. You could construct a [histogram](https://www.thoughtco.com/what-is-a-histogram-3126359) displaying the number of days that temperatures reach a certain range of values.

But all of these methods ignore a portion of the data that you have collected. Since each date is paired with the temperature reading for the day, you don‘t have to think of the data as being random. You can instead use the times given to impose a chronological order on the data.

## Application of a Time Series chart

When recording values of the same variable over an extended period of time, sometimes it is difficult to discern any trend or pattern.

1. Trend Analysis: Once the same data points are displayed graphically, some frequent patterns jump out. Time series graphs make trends easy to spot. These trends are important as they can be used to project into the future.
2. Pattern Analysis: It is important to models he business and identify cyclical patterns if any exists. The regularly increase/Decrease or up/down in the parameter with respect to time denotes possible cycle which can be identified from the graph. **This may lead to identify and predict trend. Statistical Characteristics of TS graphs**
3. **Auto correlation:** It is the similarity between observations as a function of the time lag between them.

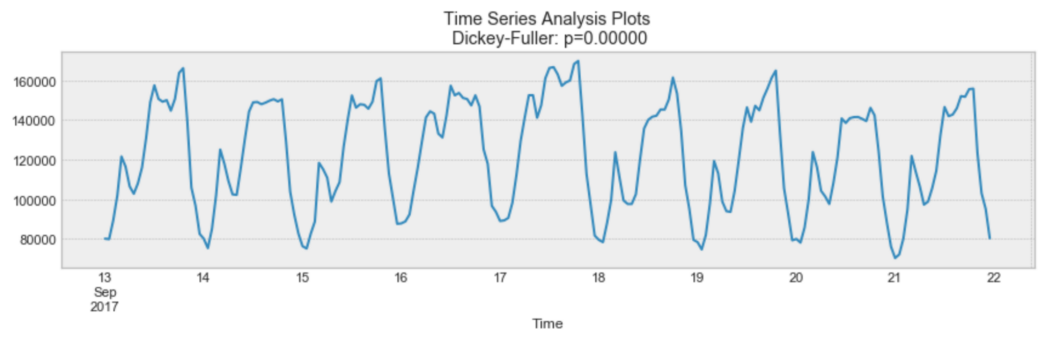


Example of an autocorrelation plot

From above graph we can realize that the first value and the 24th value have a high auto correlation. Similarly, the 12th and 36th observations are highly correlated. This means that we will find a very similar value at every 24 unit of time.

Notice how the plot looks like sinusoidal function. This is a hint for **seasonality,**and you can find its value by finding the period in the plot above, **which would give 24h**.

1. **Seasonality :** It refers to periodic fluctuations. For example, electricity consumption is high during the day and low during night, or online sales increase during Christmas before slowing down again.

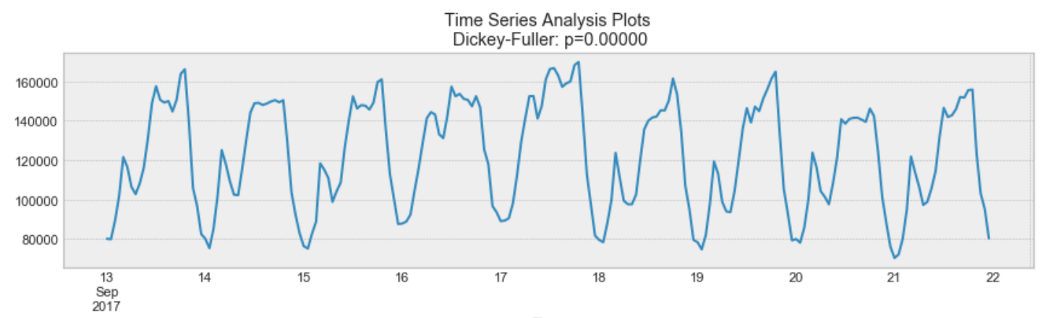


Example of seasonality

As shown in above grapg every day, you see a peak towards the evening, and the lowest points are the beginning and the end of each day which shows daily seasonality.

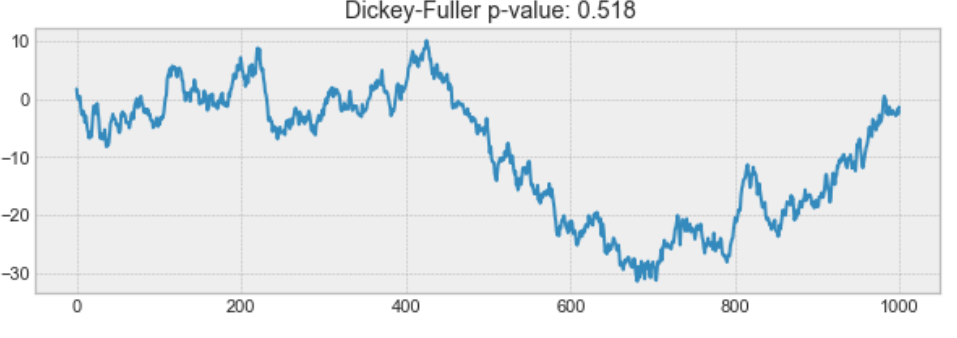
Remember that seasonality can also be derived from an autocorrelation plot if it has a sinusoidal shape. Simply look at the period, and it gives the length of the season.

1. **Stationarity:** it is an important characteristic of time series. A time series is said to be stationary if its statistical properties do not change over time. **In other words, it has constant mean and variance, and covariance is independent of time.**



Example of a stationary process

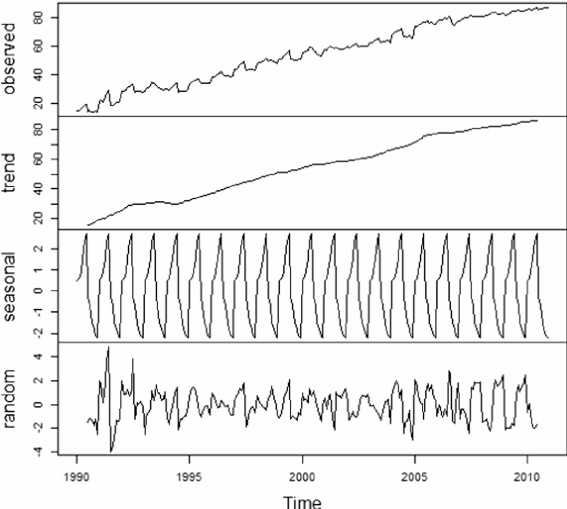
The above graph show that the process for which data is plotted is stationary. The mean and variance do not vary over time. At the same time the following graph is not containing constant mean and variance and so it is not stationary.



Note :

* Often, stock prices are not a stationary process, since we might see a growing trend, or its volatility might increase over time (meaning that variance is changing).
* Ideally, we want to have a stationary time series for modelling. Of course, not all of them are stationary, but we can make different transformations to make them stationary.

**Following figure depics the difference in trend and pattern analysis**



For the analysis we need to model the time series data before identifying trend and Patterns.

**Modeling time series**

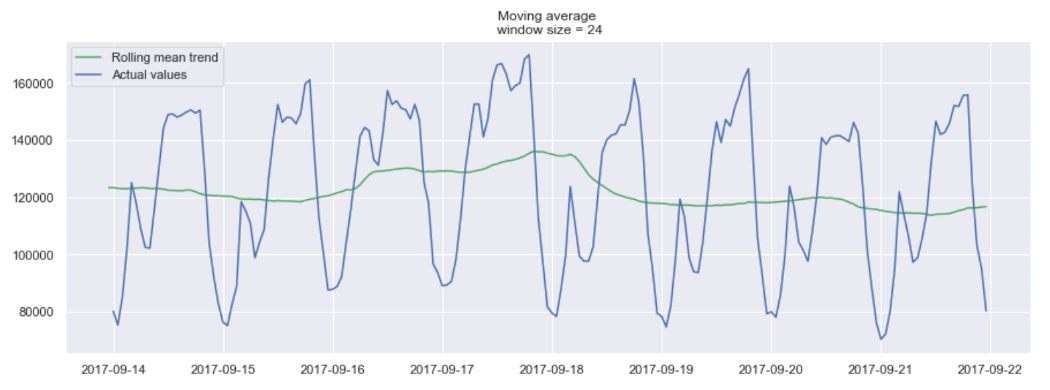
There are many ways to model a time series in order to make predictions. Here, we will cover the most famous techniques

1. **Moving average**

The moving average model is probably the most naive approach to time series modelling. This model simply states that the next observation is the mean of all past observations.

Although simple, this model might be surprisingly good and it represents a good starting point.

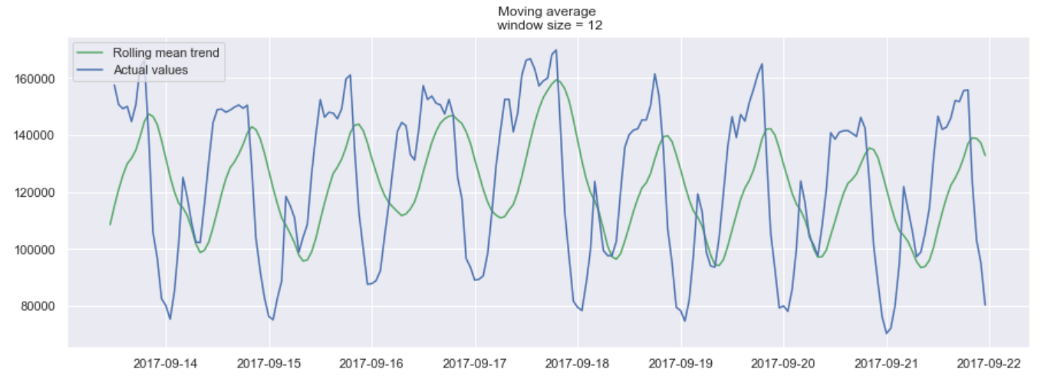
**Otherwise, the moving average can be used to identify interesting trends in the data.** We can define a window to apply the moving average model to smooth the time series, and highlight different trends.



Example of a moving average on a 24h window

In the plot above, we applied the moving average model to a 24h window. The green line smoothed the time series, and we can see that there are 2 peaks in a 24h period.

Of course, the longer the window, the smoother the trend will be. Below is an example of moving average on a smaller window.

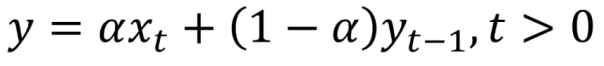


Example of a moving average on a 12h window

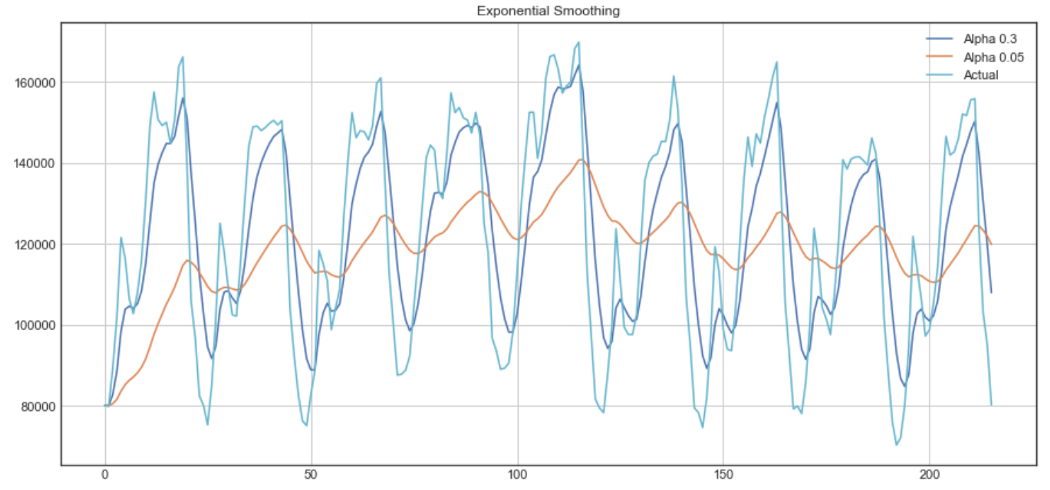
1. **Exponential smoothing**

Exponential smoothing uses a similar logic to moving average, but this time, a different decreasing weight is assigned to each observations. In other words, less importance is given to observations as we move further from the present.

Mathematically, exponential smoothing is expressed as:



Where  alpha is a smoothing factor that takes values between 0 and 1. It determines how fast the weight decreases for previous observations.



The above plot shows that the dark blue line represents the exponential smoothing of the time series using a smoothing factor of 0.3, while the orange line uses a smoothing factor of 0.05.

It also depics that, smaller smoothing factor will result in smoother time series. As smoothing factor approaches 0, we get the moving average model.